ANTI-CONCENTRATION OF INHOMOGENEOUS RANDOM WALKS

HOI H. NGUYEN

Abstract. We provide a characterization for anti-concentration of inhomogeneous random walks in non-abelian groups. In application we extend the classical bounds by Erdős-Littlewood-Offord and Sárközy-Szemerédi to non-abelian settings.

1. Introduction

Let $G = (G, \cdot)$ be an ambient group which is not necessarily abelian. Let $A_1, \dots, A_n$ be finite but not necessarily symmetric sets. Let μ_i be any probability distribution on A_i such that

$$\min_i \Big\{ \mu_i(a), a \in A_i \Big\} > p_0, \tag{1}$$

for some parameter $p_0 > 0$ which is allowed to depend on n in some cases.

We define the concentration probability of the random walk generated by $\mu_1, \dots, \mu_n$ to be

$$\rho(\mu_1, \dots, \mu_n) := \|\mu_n * \cdots * \mu_1\|_\infty = \max_{g \in G} \mu_n * \cdots * \mu_1(g).$$

Here the discrete convolution is defined as

$$\mu * \nu(g) := \sum_{h \in \mathrm{supp}(\mu)} \mu(h)\nu(h^{-1}g).$$

Thus in contrast to the classical setting of random walks, our concern here is on *inhomogeneous* ones where the supports A_i of μ_i can be totally different.

In the abelian setting with $G = \mathbf{C}$ and with $\mu_i(a_i) = \mu_i(-a_i) = 1/2$, the classical result of Erdős [6] and Littlewood-Offord [9] shows

Theorem 1.1 (forward Erdős-Littlewood-Offord)**.** *Assume that a_i are all non-zero complex numbers, then*

The author is partly supported by research grant DMS-1600782.

$$\rho(\mu_1, \dots, \mu_n) \leq \frac{\binom{n}{\lfloor n/2 \rfloor}}{2^n}. \quad (2)$$

This result was improved later by Sárközy and Szemerédi [13] (see also [7, 14, 10]) under an extra assumption.

Theorem 1.2 (forward Sárközy-Szemerédi)**.** *Assume that a_i are distinct complex numbers, then*

$$\rho(\mu_1, \dots, \mu_n) \leq (\sqrt{\frac{24}{\pi}} + o(1)) \frac{1}{n^{3/2}}. \quad (3)$$

All of these results are optimal. We also refer the reader to the work by Halász [8], and to the survey [12] for further extensions and applications of these results.

1.3. **Non-abelian results.** Although in the abelian setting the a_i can be different, the ordering of the random steps does not matter. This is also the case for classical random walks (in either abelian or non-abelian groups) of the form $\mu * \cdots * \mu$. However, this pleasant property totally breaks down for inhomogeneous random walks in non-abelian groups, and this makes the analysis quite intractable.

As far as we are concerned, not much is known in the general non-abelian setting for inhomogeneous random walks. One related result we could find in the literature is from Varopolous' book [5, Chapter VII 1.2.] where G is a unimodular compactly generated group with polynomial volume growth of order D, and where the inhomogeneous random walk is generated by the μ_i of uniformly bounded density functions. It was shown in this case that the density function of $\mu_n * \cdots * \mu_1$ is bounded from above by $n^{-D/2}$; we refer the reader to [5] for more details.

Another result, which is directly relevant to our study, is a recent work by Pham and Vu [15, Theorem 1.3].

Theorem 1.4 (forward Erdős-Littlewood-Offord for matrices)**.** *Let m, n, s be integers and let $a_i, 1 \leq i \leq n$ be elements of $GL_m(\mathbf{C})$ with order at least s. Assume that $A_i = \{a_i, a_i^{-1}\}$ and $\mu_i(a_i) = \mu_i(a_i^{-1}) = 1/2$. Then*

$$\rho(\mu_1, \dots, \mu_n) \leq 141 \max\{\frac{1}{s}, \frac{1}{n^{1/2}}\}.$$

This bound is optimal up to the explicit multiplicative constant.

One of the main goals of this note is to show the following analog of Theorem 1.4 in asymptotic form.

Theorem 1.5 (forward Erdős-Littlewood-Offord for general groups)**.** *For any $\delta > 0$, there exist n_0 and $0 < \varepsilon < 1$ such that the following holds for $n \geq n_0$. Assume that the distributions μ_i in G satisfy (1) with $p_0 \geq n^{-\varepsilon^3}$ and such that each A_i contains a pair of elements a_i, a_i' with $a_i a_i'^{-1}$ being order of at least s, then*

$$\rho(\mu_1,\dots,\mu_n) \leq \max\{\frac{1}{s}, \frac{1}{n^{1/2-\delta}}\}.$$

In particularly, assume that $a_1,\dots,a_n$ are of order at least s in G and the supports A_i contain $\{a_i, a_i^{-1}\}$, then the same conclusion holds.

Next, motivated by Sárközy-Szemerédi's result, one might also be interested in getting a non-trivial bound for $\rho(\mu_1,\dots,\mu_n)$ when the μ_i are essentially different. Our next result shows

Theorem 1.6 (forward Sárközy-Szemerédi for general groups)**.** *For any $\delta > 0$, there exist n_0 and $0 < \varepsilon < 1$ such that the following holds for $n \geq n_0$. Assume that the distributions μ_i in G satisfy (1) with $p_0 \geq n^{-\varepsilon^3}$ such that each A_i contains a pair of elements a_i, a_i' so that $a_i a_i'^{-1}, 1 \leq i \leq n$, are all distinct. Then*

$$\rho(\mu_1,\dots,\mu_n) \leq \frac{1}{n^{1-\delta}}.$$

In particularly, assume that $a_1,\dots,a_n$ are n distinct elements of G and the supports A_i contain $\{id_G, a_i\}$, then the same conclusion holds.

In general the bound $n^{-1+o(1)}$ above is asymptotically sharp by the example that a_i are elements of subgroups of $\Theta(n)$ elements. We also note that the conclusion fails in general if A_i contains $\{a_i^{-1}, a_i\}$ instead of $\{id_G, a_i\}$ because the a_i might have order two.

1.7. **Method of proof.** The way we prove Theorem 1.5 and Theorem 1.6 has its origin in [21]. Notably, we will not be working directly with *forward* results as in Theorem 1.1,1.2,1.4 but with the *backward* ones. Roughly speaking, say if we want to prove Theorem 1.5, assume for contradiction that

$$\rho(\mu_1,\dots,\mu_n) \gg \max\{\frac{1}{s}, \frac{1}{n^{1/2-\delta}}\}. \tag{4}$$

We then show that there exists a support A_i of μ_i where $a_i a_i'^{-1}$ is of order smaller than s, this violates the assumption of the theorem.

Similarly, to prove Theorem 1.6 we assume for contradiction that

$$\rho(\mu_1,\dots,\mu_n) \gg \frac{1}{n^{1-\delta}}. \tag{5}$$

Then there exists a set of size $o(n^{1/2})$ that contains most of the $a_i a_i'^{-1}$, which again contradicts our assumption.

The study of (4) and (5), in its general framework, is called the *inverse Littlewood-Offord problem*. This was raised by Tao and Vu [19, 20] about ten years ago.

Problem 1.8. *Characterize the sets $A_1, \dots, A_n$ when*

$$\rho(\mu_1, \dots, \mu_n) \geq n^{-O(1)}.$$

We will devote the rest of this section to discuss this problem. To give an example of sets of large concentration probability, we first introduce some arithmetic structures.

Definition 1.9 (progression)**.** Let $u_1, \dots, u_r$ be elements of G, and let $(N_1, \dots, N_r)$ be a vector of positive integers. Then the set of all products in the u_i and their inverses in which each u_i and its inverse appear at most N_i times is called a *progression of rank r and size lengths $N_1, \dots, N_r$*, and is denoted by $P(u_1, \dots, u_r; N_1, \dots, N_r)$ (or P for short).

When G is abelian, it is not hard to see that progressions grow very slow under addition in G. Thus if $A_1, \dots, A_n \subset P(u_1, \dots, u_r; N_1, \dots, N_r)$ with $\prod_i N_i = n^{O(1)}$ then $\rho(\mu_1, \dots, \mu_n) \geq n^{-O(1)}$. It was shown by Tao and Vu in [20, 21] (see also [11] and [17]) that the converse is also true.

Theorem 1.10 (inverse Erdős-Littlewood-Offord)**.** *Let G be a torsion-free abelian group. Let $A > 0$ and $1 > \varepsilon > 0$ be given constants, and let m be any quantity between n^{ε} and $n^{1-\varepsilon}$. Assume that $\mu_i(a_i) = \mu_i(-a_i) = 1/2$ and*

$$\rho(\mu_1, \dots, \mu_n) \geq n^{-A}.$$

Then there exists a symmetric progression $P = P(u_1, \dots, u_r; N_1, \dots, N_r)$ of rank $r = O(1)$ and size $O(\rho^{-1}/m^{r/2})$ and there exist $n - m$ indices $i \in [1, n]$ such that

$$A_i \subset P.$$

Our method develops a non-abelian counterpart of this result. We remark that the recent work by Tao [17], among other things, studies the distribution μ when $\rho(\mu, \dots, \mu) \geq n^{-A}$. The abelian inverse result, Theorem 1.10 above, can be viewed as a very special case of the general framework of [17], but the results there do not seem to directly cover our current setting of inhomogeneous random walks.

Notice that for general G, one does not expect the condition $A_i \subset P$ to imply the largeness of $\rho(\mu_1, \dots, \mu_n)$. However, it would do if we know that the progressions P are "almost abelian".

Definition 1.11 (nilprogression and coset nilprogression, [4])**.** Suppose that G is a group and $r \geq 1, s \geq 0$ are integers.

- A *nilprogression* of rank r and step s is a progression $P(u_1, \dots, u_r; N_1, \dots, N_r)$ with the property that every iterated commutator of degree $s + 1$ in the generators $u_1, \dots, u_r$ equals the identity id_G.

- A *coset nilprogression* of rank r and step s is a set of the form $\pi^{-1}(P)$, where P is a nilprogression of rank r and step s in a quotient group G_0/H, where H is a finite normal subgroup of a subgroup G_0 of G and $\pi : G_0 \to G_0/H$ is the quotient map.

Thus coset nilprogressions can be written under the form HP, where H is a finite subgroup which commutes as set with elements of the subgroup $\langle P \rangle$ generated by P.

We next introduce a special type of nilprogression.

Definition 1.12 (C-normal form, [4])**.** Let $C \geq 1$. A nilprogression $P(u_1, \dots, u_r; N_1, \dots, N_r)$ is said to be in C-normal form if the following axioms are obeyed.

- (Upper triangular form) For every i, j with $1 \leq i < j \leq r$ and for all four choices of signs for the commutators
$$[u_i^{\pm 1}, u_j^{\pm 1}] \in P(u_{j+1}, \dots, u_r; \frac{CN_{j+1}}{N_i N_j}, \dots, \frac{CN_r}{N_i N_j}).$$
- (Local properness) The expressions $u_1^{n_1} \cdots u_r^{n_r}$ are distinct as $n_1, \dots, n_r$ range over integers with
$$|n_i| \leq \frac{1}{C} N_i.$$
- (Volume bound) One has
$$\frac{1}{C}(2\lfloor N_1 \rfloor + 1) \dots (2\lfloor N_r \rfloor + 1) \leq |P| \leq C(2\lfloor N_1 \rfloor + 1) \dots (2\lfloor N_r \rfloor + 1).$$

A coset nilprogression $\pi^{-1}(P)$ is said to be in C-normal form if the nilprogression P is C-normal in the quotient group G_0/H.

We also refer the reader to [2, 22] for several asymptotic equivalence between progressions and nilprogressions in nilpotent groups. A crucial property of coset nilprogressions in C-normal form is that their products grow polynomially slow (see [4, Proposition C.5]),

$$|HP^n| = n^{O_{C,r}(1)}|H||P|. \tag{6}$$

As such, similarly to the abelian case, coset nilprogressions are examples of sets of high concentration probability.

Example 1.13. *Assume that HP is a nilprogression in C-normal form with rank r and step s of order $O(1)$, and with small cardinality $|HP| = n^{O(1)}$.*

- *Assume that $A_1, \dots, A_n \subset HP$, then by* (6) *and by the pigeonhole principle,*
$$\rho(\mu_1, \dots, \mu_n) \geq n^{-O(1)}.$$

- *More generally, assume that there is a finite set X with $|X| = O(1)$ such that for each $1 \leq i \leq n$ and each $a \in A_i$ there exists a permutation $\sigma_a \in Sym(X)$ such that for all $x \in X$,*

$$a \in xHP(\sigma_a(x))^{-1}.$$

It is clear that in this case

$$A_n \dots A_1 \subset XHP \dots HPX^{-1} = XHP^n X^{-1}.$$

Hence $|A_n \dots A_1| = n^{O(1)}$, and so by the pigeon principle

$$\rho(\mu_1, \dots, \mu_n) \geq n^{-O(1)}.$$

By adapting the method of [17], we will show the converse of the above.

Theorem 1.14. *Let G be a non-abelian group. Let $A > 0$ an $1 > \varepsilon > 0$ be given constants. Assume that the distributions μ_i in G satisfy (1) with $p_0 \geq n^{-\varepsilon^3}$ and such that*

$$\rho = \rho(\mu_1, \dots, \mu_n) \geq n^{-A}.$$

Then there exists a coset nilprogression HP with the following properties.

(1) P is in C-normal form with $C = O(1)$ and with rank and step $r, s = O(1)$,

(2) $|HP| = O(\rho^{-1})$,

(3) there is a finite set X of cardinality $|X| = O(1)$, and consecutive indices $i_0, \dots, i_0 + n'$ with $n' = n^{1-O(\varepsilon)}$ such that the following holds: for each $a, a' \in A_i, i_0 \leq i \leq i_0 + n'$ there exists a permutation $\sigma \in Sym(X)$ such that for all $x \in X$,

$$aa'^{-1} \in xHP(\sigma(x))^{-1}.$$

Here the implied constants depend on ε and A but not on G.

The bounds for p_0 and n' above can be slightly improved but our final conclusion is not optimal, we refer the reader to Conjecture 4.3 for a possible extension of this theorem. Although our characterization captures only n' consecutive μ_i with some $n' = n^{1-O(\varepsilon)}$ (in comparison to $n' = (1 - o(1))n$ from Theorem 1.10), we can certainly run the argument at other segments; the obtained information is usually sufficient for asymptotic estimates.

Theorem 1.14 heuristically supports the phenomenon that for the type of inhomogeneous random walks under consideration it is not *at all coincident* when the concentration probability is polynomially large at some sufficiently large step n. Indeed, *generic* inhomogeneous random walks should have *extremely small* concentration probability. To illustrate this point furthermore, allow us to give an example in the simple context of $\mathbf{Sl}_2(\mathbf{R})$ in connection to the discrete Anderson-Bernoulli model in 1D. The result is by no mean important but we are not able to find similar statement in the literature.

Consider the random walk generated by transfer matrices $g_i := \begin{pmatrix} E+\lambda\varepsilon_i & -1 \\ 1 & 0 \end{pmatrix}$ where $E, \lambda \in \mathbf{R}, \lambda > 0$ are given parameters, and where $\varepsilon_i, 1 \leq i \leq n$ are independent random variables with possibly different discrete distributions μ_i in $\mathbf{R}$ satisfying (1). Assume furthermore that for any collection of $n^{1-\varepsilon}$ consecutive distributions $\mu_{i_0}, \dots, \mu_{i_0+n^{1-\varepsilon}}$ there is a distribution μ_i whose support contains a symmetric pair $\{a_i, -a_i\}$ with a_i is greater than a given positive parameter γ.

Theorem 1.15. *Let be given $E, 0 < \lambda, 0 < \gamma$ and $0 < \varepsilon < 1$, the following holds for $\mu_1, \dots, \mu_n$ satisfying the above conditions with sufficiently large n depending on λ, γ and ε*

$$\sup_{g \in \mathbf{Sl}_2(\mathbf{R})} \mathbf{P}(g_1 \dots g_1 = g) = n^{-\omega(1)}.$$

It is possible that the bound in Theorem 1.15 is sub-exponential or even smaller, but we are unable to confirm this. Let us now discuss the proof of Theorem 1.14. To ease the presentation, we will decompose the proof into three parts.

(1) In the first step we will rely on the celebrated result by Breuillard, Green and Tao [4] to obtain structures in the supports of large convolution sequences $\mu_{i_0+2l_0^*} * \cdots * \mu_{i_0+l_0^*}$ and $\mu_{i_0+l_0^*} * \cdots * \mu_{i_0}$ for some $l_0^* = n^{1-o(1)}$ and $i_0 = o(n)$. To arrive at the point of applying [4], we will use a simple dyadic argument and an asymmetric version of Balog-Szemerédi-Gowers theorem due to Tao [16].

(2) In the second step, by following the mentioned work by Tao [17], we obtain structures in the supports of smaller convolution sequences of type $\mu_{i_0+l_0^*} * \cdots * \mu_{i_0+i}, 0 \leq i \leq l_0$. The main focus of this step is on a semi-metric defined with respect to the structures obtained in Step 1.

(3) In the last step, we improve upon Step 2 to obtain structures in the support of each individual μ_{i_0+i}.

As we can see, our proof of 1.14 mainly relies on [4] and [17], so the implicit constants of this result, and hence of Theorem 1.5, Theorem 1.6 and Theorem 1.15, are ineffective.

Notation. Throughout this paper, n as an asymptotic parameter going to infinity. We write $X = O_K(Y)$, $X \ll_K Y$, or $Y \gg_K X$ to denote the claim that $|X| \leq CY$ for some constant C that depends on K. We also use $o(Y)$ to denote any quantity bounded in magnitude by $c(n)Y$ for some $c(n)$ that goes to zero as $n \to \infty$. Again, the function $c(.)$ is permitted to depend on fixed quantities.

The rest of the note is organized as follows. The proof of Theorem 1.14 is presented in Sections 2, 3 and 4. Theorem 1.5 will be shown in Section 5 by following the same ideas with some modifications. Finally, the proof of Theorem 1.6 and Theorem 1.15 will be presented in Section 6 and Section 7 respectively.

2. Proof of Theorem 1.14: first step

The main result of this section is Theorem 2.7. First of all, we introduce some elementary inequalities to be used.

Claim 2.1 (Young's inequality)**.** *Let μ and ν be probability measures with finite support in G. Then*

- $$\|\mu\|_\infty \le \|\mu\|_2 \le \|\mu\|_\infty^{1/2};$$
- $$\|\mu * \nu\|_\infty \le \|\mu\|_2 \|\nu\|_2;$$
- $$\|\mu * \nu\|_2 \le \min\{\|\mu\|_2, \|\nu\|_2\}.$$

Because of the second inequality, by passing to a subsequence of size at least $n/2$ when needed, instead of assuming $\|\mu_n * \cdots * \mu_1\|_\infty \ge \rho$, we assume that

$$\|\mu_n * \cdots * \mu_1\|_2^2 \ge \rho. \tag{7}$$

For short, for $i < j$ we write

$$\mu_{[i,j]} := \mu_j * \cdots * \mu_i.$$

Claim 2.2. *Let $0 < \varepsilon < 1$ be given. There exist i_0, l_0 with $i_0 + 4l_0 \le n$ and $l_0 \ge n^{1-\varepsilon/2}$ such that*

$$\|\mu_{[i_0, i_0+4l_0]}\|_2 \ge c \max_{i_0 \le i \le i_0 + 7l_0/2} \|\mu_{[i, i+l_0/2]}\|_2, \tag{8}$$

where c is a sufficiently small constant depending on ε.

Proof. (of Claim 2.2) The proof is standard. Assume otherwise, then we can find a nested sequence $[1,n] \supset [i_1, i_1 + 4l_1] \supset [i_2, 4l_2] \supset \cdots \supset [i_k, i_k + 4l_k]$ such that $l_{j+1} = l_j/8$ and that

$$\|\mu_{[i_j, i_j + 4l_j]}\|_2 \le c \|\mu_{[i_{j+1}, i_{j+1} + 4l_{j+1}]}\|_2.$$

However, as $\|\mu_{[1,n]}\|_2 \ge n^{-O(1)}$ and $\|\mu_{[.]}\|_2 \le 1$, the nested sequence above must have at most $k = O(\log_{1/c} n)$ terms. By definition

$$l_k = \Omega(\frac{n}{8^k}) = \Omega(n^{1-\varepsilon/2}).$$

□

With i_0, l_0 from Claim 2.2 we have

$$\prod_{0\le j\le n^{\varepsilon/2}/2} \frac{\|\mu_{[i_0+l_0-(j+1)n^{1-\varepsilon},i_0+2l_0]}\|_2}{\|\mu_{[i_0+l_0-jn^{1-\varepsilon},i_0+2l_0]}\|_2} = \frac{\|\mu_{[i_0+l_0-(n^{\varepsilon/2}/2+1)n^{1-\varepsilon},i_0+2l_0]}\|_2}{\|\mu_{[i_0+l_0,i_0+2l_0]}\|_2} \ge \frac{\|\mu_{[i_0,i_0+2l_0}\|_2}{\|\mu_{[i_0,i_0+l_0]}\|_2} \ge c.$$

Thus there exists $0 \le j \le n^{\varepsilon/2}/2$ such that

$$\frac{\|\mu_{[i_0+l_0-(j+1)n^{1-\varepsilon},i_0+2l_0}\|_2}{\|\mu_{[i_0+l_0-jn^{1-\varepsilon},i_0+2l_0]}\|_2} \ge 1 - n^{-\varepsilon}. \tag{9}$$

Set $j_0 := i_0 + l_0 - jn^{1-\varepsilon}$ and $l_0^* := i_0 + 2l_0 - j_0$. Then

$$l_0^* \ge l_0 - n^{1-\varepsilon/2}/2 \ge l_0/2.$$

Combine Claim 2.2 and (9), using the third monotonicity property from Claim 2.1 we obtain

Lemma 2.3. *The exist j_0, l_0^* with $l_0^* \ge n^{1-\varepsilon}$ such that*

$$\|\mu_{[j_0-l_0^*,j_0+l_0^*]}\|_2 \ge c \max\Big\{\|\mu_{[j_0-l_0^*,j_0-1]}\|_2, \|\mu_{[j_0,j_0+l_0^*]}\|_2\Big\} \tag{10}$$

and

$$|\mu_{[j_0-m,j_0+l_0^*]}\|_2 \ge (1-n^{-\varepsilon})\|\mu_{[j_0,j_0+l_0^*]}\|_2 \text{ for all } m \le n^{1-\varepsilon}. \tag{11}$$

Note that although we vary m in (11), the inequality is clearly most meaningful at $m = n^{1-\varepsilon}$. For the rest of this section, we will focus on (10). For brevity, write

$$\mu := \mu_{[j_0,j_0+l_0^*]}, \text{ and } \nu := \mu_{[j_0-l_0^*,j_0-1]}.$$

We can rewrite (10) to

$$c \max\Big\{\|\mu\|_2, \|\nu\|_2\Big\} \le \|\mu * \nu\|_2 \le \min\Big\{\|\mu\|_2, \|\nu\|_2\Big\}. \tag{12}$$

To exploit this nice property, we will need an important notion of approximate group.

Definition 2.4. [4, Definition 1.2] Let $K \ge 1$. A K-approximate group in a group G is a multiplicative set A with the following properties

- the set A is symmetric: $id_G \in A$ and $a^{-1} \in A$ if $a \in A$;
- there is a symmetric subset $X \subset A^3$ with $|X| \le K$ such that
$$AA \subset XA.$$

By using the asymmetric weighted Balog-Szemerédi-Gowers theorem we obtain the following analog of [3, Proposition A.1].

Lemma 2.5. *Assume that μ and ν are probability measures such that*

$$\|\mu * \nu\|_2 \geq \frac{1}{K} \max\{\|\mu\|_2, \|\nu\|_2\}.$$

Then there is a $O(K^{O(1)})$-approximate subgroup A of G and $x_0, y_0 \in G$ such that

$$|A| \ll K^{O(1)} (\max\{\|\mu\|_2, \|\nu\|_2\})^{-2}.$$

and

$$\mu(x_0 A), \nu(A y_0) \gg K^{-O(1)}.$$

In application, as by (12) we will set

$$K := c^{-1}.$$

Proof. (of Lemma 2.5) We apply the machinery from [3, Proposition A.1] and [16, Theorem 4.6]. Set

$$\delta := \frac{1}{100K^2} \text{ and } M = 10K.$$

Define

$$\mu' := \mu 1_{\mu \geq M\|\mu\|_2^2}, \mu'' := \mu 1_{\mu \leq \delta \|\mu\|_2^2}, \text{ and } \tilde{\mu} := \mu - \mu' - \mu''.$$

We note that

$$\sum_{g \in \text{supp}(\mu)} \mu'(g) \leq \sum_{g \in \text{supp}(\mu)} \mu'(g) \frac{\mu'(g)}{M\|\mu\|_2^2} \leq \frac{1}{10K}.$$

Furthermore,

$$\sum_{g \in \text{supp}(\mu)} \mu''(g)^2 = \sum_{g \in \text{supp}(\mu)} \mu(g)^2 1_{\mu \leq \delta \|\mu\|_2^2} \leq \delta \|\mu\|_2^2.$$

As such, by Young's inequality,

$$\|\mu' * \nu\|_2 \leq \min\Big\{\|\mu'\|_2\|\nu\|_1, \|\mu'\|_1\|\nu\|_2\Big\} \leq \frac{1}{10K}\|\nu\|_2 \leq \frac{1}{10}\|\mu * \nu\|_2,$$

and

$$\|\mu'' * \nu\|_2 \leq \min\Big\{\|\mu''\|_2\|\nu\|_1, \|\mu''\|_1\|\nu\|_2\Big\} \leq \delta^{1/2}\|\mu\|_2 \leq \frac{1}{10}\|\mu * \nu\|_2.$$

Thus by the triangle inequality, $\|\tilde{\mu}\|_2$ and $\|\tilde{\mu} * \nu\|_2$ are comparable to $\|\mu * \nu\|_2$,

$$\|\mu\|_2 \geq \|\tilde{\mu}\|_2 \geq \|\tilde{\mu} * \nu\|_2 \geq \frac{4}{5}\|\mu * \nu\|_2 \geq \frac{4c}{5}\|\mu\|_2.$$

By doing similarly with ν, we obtain

$$\|\tilde{\mu} * \tilde{\nu}\|_2 \geq \frac{1}{2K}\max\{\|\tilde{\mu}\|_2, \|\tilde{\nu}\|_2\}. \tag{13}$$

Setting $B_1 := \mathrm{supp}(\tilde{\mu}), B_2 := \mathrm{supp}(\tilde{\nu})$. Then by definition of $\tilde{\mu}$ and $\tilde{\nu}$

$$|B_1|, |B_2| \asymp_K \|\mu\|_2^{-2} \text{ and } \mu(B_1), \nu(B_2) \asymp_K 1.$$

Also, by (13)

$$E(B_1, B_2) \gg_K |B_1|^3,$$

where the implicit constants depend polynomially on K, and where $E(B_1, B_2)$ is the multiplicative energy,

$$E(B_1, B_2) := \#\Big\{(b_1, b_1', b_2, b_2') \in (B_1^2 \times B_2^2) : b_1 b_2 = b_1' b_2'\Big\}.$$

By Tao's result on product set estimates for non-commutative groups [16, Theorem 5.2], there exist subsets $B_1' \subset B_1, B_2' \subset B_2$ with $|B_i'| \gg \frac{|B_i|}{K^{O(1)}}$ and such that

$$|B_1' B_2'| \leq K^{O(1)}|B_1|.$$

Also by [16, Theorem 4.6], there exists a $O(K^{O(1)})$-approximate group A of size $O(K^{O(1)}|B_1|)$ and a finite set Y of cardinality $O(K^{O(1)})$ such that

$$B_1' \subset YA \text{ and } B_2' \subset AY.$$

Thus there exists $x_0 \in Y$ and $y_0 \in Y$ such that

$$|B_1' \cap x_0 A|, |B_2' \cap A y_0| \geq |B_1|/O(K^{O(1)}) \geq \|\mu\|_2^{-2}/O(K^{O(1)}).$$

This completes the proof of Lemma 2.5. □

Our next ingredient is a simplified version [1] of the mentioned celebrated result by Breuillard, Green, and Tao.

Theorem 2.6. [4, Theorem 2.10] *Let A be a finite K-approximate group in a global group G. Then A^4 contains a coset nilprogression HP of rank and step $O_K(1)$ and $|P| \gg_K |A|$. Furthermore, P can be taken to be in $O_K(1)$-normal form.*

Combine Lemma 2.5 with $K = c^{-1}$ and Theorem 2.6, after a covering argument (as $A^4 \subset X^3 A$ for approximate group A), we obtain the following.

Theorem 2.7. *Assume as in Lemma 2.5, then there exists a coset nilprogression HP (in $O_c(1)$-normal form) with $|HP| \ll_c \|\mu\|_2^{-2}$ and there exist $x_0, y_0 \in G$ such that*

$$\mu(x_0 HP), \nu(HP y_0) \gg_c 1.$$

In particularly, Theorem 2.7 holds for μ and ν defined after Lemma 2.3.

For later steps, it will be more convenient to pass to a sub nilprogression Q of P which is slightly more "proper". Let $D \leq 1/\varepsilon$ be a constant to be chosen sufficiently large depending on other parameters (such as rank, step, C-normal form) of the structure P obtained in Theorem 2.7. Consider the nillprogression

$$Q := P_{1/CD^2} = P(u_1, \dots, u_r; M_1, \dots, M_r) \text{ with } M_i = \frac{1}{CD^2} N_i. \tag{14}$$

By definition, the following holds for Q

(i) for every $1 \leq i < j \leq r$,

$$[u_i^{\pm 1}, u_j^{\pm 1}] \in P(u_{j+1}, \dots, u_r; \frac{M_{j+1}}{D^2 M_i M_j}, \dots, \frac{M_r}{D^2 M_i M_j});$$

(ii) the expressions $u_1^{k_1} \cdots u_r^{k_r}$ are distinct for all $k_1, \dots, k_r$ with

$$|k_i| \leq D M_i;$$

(iii)

$$|HQ| \asymp_c |HP|.$$

[1] This result holds in more general setting where G can be local, see [4].

We show that if $x \in HQ$ and $x^2 \in HQ$ then x is asymptotically an element of $HQ_{1/2}$.

Claim 2.8. *The following holds with D sufficiently large*

(1) Assume that x is an element of Q where each u_i and u_i^{-1} appears with frequencies n_i, n_i' respectively. Then u can be written as $x = u_1^{m_1-m_1'} \dots u_n^{m_r-m_r'}$ with

$$\max\{|m_i - n_i|, |m_i' - n_i'|\} \leq \frac{M_i}{D} \text{ for all } i.$$

(2) Assume that $x = u_1^{n_1} \dots u_n^{n_r} \in Q$, then $x^2 = u_1^{m_1} \dots u_r^{m_2}$ with

$$|m_i - 2n_i| \leq \frac{M_i}{D}.$$

(3) Assume that $x \in HQ$ and such that $x^2 \in HQ$, then

$$x \in HQ_{(1+\frac{1}{D})}.$$

We insert here a proof for completion.

Proof. (of Claim 2.8) We will prove the first assertion, the second one follows similarly. Assume that in the representation of x there are exactly $n_i^{(0)} = n_i$ and $n_i'^{(0)} = n_i'$ copies of u_i and u_i^{-1} respectively for $1 \leq i \leq r$. We will move all copies of u_1 and u_1^{-1} to the left. By (i), each step of replacing of $u_i u_1$ by $u_1 u_i [u_i, u_1]$ would change the multiplicities $n_j^{(0)}$ of u_j to $n_j^{(1)}$ where $i+1 \leq j \leq n$ and

$$|n_j^{(1)} - n_j^{(0)}| \leq \frac{M_j}{D^2 M_i M_1}.$$

Thus, after moving the first copy of u_1 all the way to the left after some $k_1 \leq 2(M_1 + \dots + M_r)$ replacements, one has

$$|n_j^{(k_1)} - n_j^{(0)}| \leq \frac{M_j}{D^2 M_1} \sum_{i \leq j-1} \frac{s_i^{(0)}}{M_i} \leq \frac{r M_j}{D^2 M_1}, \tag{15}$$

where we used the fact that $s_i^{(0)}$, the number of times u_1 meets u_i, is bounded by $s_i^{(0)} \leq n_i^{(0)} \leq M_i$. As a consequence, after n_i steps of moving all u_1 to the left, one has [2]

$$|n_j^{(k_1 + \dots + k_{n_1})} - n_j^{(0)}| \leq n_j^{(0)} \cdot O(\frac{r M_j}{D^2 M_1}) = O(\frac{r M_j}{D^2}). \tag{16}$$

[2]Strictly speaking, the bounds of $s_i^{(0)}$ from (15) will increase after each round of moving a copy u_1 all the way to the left, but this change is negligible.

Hence if we write $x = u_1^{n_1 - n_1'} y$, then in y the $u_j, 2 \leq j \leq r$ appears with total frequency m_j where

$$n_j - O(rM_i/D^2) \leq m_j \leq n_j + O(rM_i/D^2).$$

We apply the collecting process again for y. The process terminates after $2r$ iterations, and at the end we obtain the desired bounds assuming D to be large compared to r.

Now we show the third claim for the case of nilprogression. We write x in the form $u_1^{n_1} \dots u_n^{n_r}$ with $|n_i| \leq (1 + 1/D)M_i$ as in (1). By the second assertion,

$$x^2 = u_1^{m_1} \dots u_r^{m_r}$$

with $|m_i - 2n_i| \leq M_i/D$.

However, as the elements $u_1^{k_1} \dots u_r^{k_r}$ are distinct for all $|k_1| \leq DM_1, \dots, |k_r| \leq DM_r$, and as $x^2 \in Q$, by (1) we must have $|m_i| \leq (1 + 1/D)M_i$, and so

$$2|n_i| - M_i/D \leq (1 + 1/D)M_i.$$

Thus

$$|n_i| \leq \frac{1}{2}(1 + \frac{1}{D})M_i.$$

For the coset nilprogression case, note that if $x \in HQ$ and $x^2 \in HQ$ then $\pi(x) \in Q$ and $\pi^2(x) = \pi(x^2) \in Q$. We then argue as above for $\pi(x)$. □

By using covering arguments, one sees that Theorem 2.7 remains valid when P is replaced by Q (although with slightly worse constants). Without loss of generality we will assume our nilprogression P to satisfy Claim 2.8 from now on.

3. Proof of Theorem 1.14: second step

We next continue our proof of Theorem 1.14 by exploiting Theorem 2.7 and equation (11) from Lemma 2.3. Our main result of this section, Lemma 3.5, is obtained by following the approach of [17].

Set

$$n_0 := n^{\varepsilon}.$$

For a given coset nilprogression $HP(x_1,\dots,x_r;N_1,\dots,N_r)$, and for $g\in\langle HP\rangle$, we define the norm of g with respect to HP to be

$$\|g\|_{HP}:=\inf\Big\{\lambda : g\in HP(x_1,\dots,x_r;\lambda N_1,\dots,\lambda N_r)\Big\}.$$

For short, we will denote $HP(x_1,\dots,x_r;\lambda N_1,\dots,\lambda N_r)$ by HP_λ. Note that in the special case that $\|g\|_{HP}<1/\max\{N_1,\dots,N_r\}$ then $g\in H$.

Next, for $g\in X\langle HP\rangle X^{-1}$ we also define the norm of g with respect to X and HP as

$$\|g\|_{HP,X}:=\inf\Big\{\lambda : \exists\sigma\in Sym(X) \text{ so that } \forall x\in X, g\in xHP_\lambda(\sigma(x))^{-1}\Big\}. \tag{17}$$

Again, in the special case that $\|g\|_{HP,X}<1/\max\{N_1,\dots,N_r\}$ then there exists $\sigma\in Sym(X)$ so that for all $x\in X, g\in xH(\sigma(x))^{-1}$.

Recall from the second property (11) of Lemma 2.3 that

$$\|\mu*\eta_m\|_2\geq(1-1/n_0)\|\mu\|_2$$

with $\mu=\mu_{[j_0,j_0+l_0^*]}$ and $\eta_m=\mu_{[j_0-m,j_0-1]}$ for any $m\leq n^{1-\varepsilon}$.

This can be rewritten as

$$\int_G\int_G\|\mu*\delta_g-\mu*\delta_h\|_2^2 d\eta_m(g)d\eta_m(h)=2(\|\mu\|_2^2-\|\mu*\eta_m\|_2^2)\leq\frac{4}{n_0}\|\mu\|_2^2.$$

Motivated by this, we call a pair $(g,h)\in G^2$ in $\mathrm{supp}(\eta_m)\times\mathrm{supp}(\eta_m)$ *typical* if

$$\|\mu*\delta_g-\mu*\delta_h\|_2\leq\frac{1}{n_0^{1/2-\varepsilon/2}}\|\mu\|_2. \tag{18}$$

Note that η_m has discrete support. Let T_{η_m} denote the set of typical pairs.

Claim 3.1. *For η_m-asymptotically almost surely, any pair $(g,h)\in T_{\eta_m}$ is typical. More precisely,*

$$\sum_{(g,h)\notin T_{\eta_m}}\eta_m(g)\eta_m(h)\leq\frac{4}{n_0^{\varepsilon}}.$$

Proof. (of Claim 3.1) By definition,

$$\frac{1}{n_0^{1-\varepsilon}}\|\mu\|_2^2 \sum_{(g,h)\notin T_{\eta_m}} \eta_m(g)\eta_m(h) \le \sum_{(g,h)\notin T_{\eta_m}} \|\mu * \delta_g - \mu * \delta_h\|_2^2 \eta_m(g)\eta_m(h) \le \frac{4}{n_0}\|\mu\|_2^2.$$

Thus

$$\sum_{(g,h)\notin T_{\eta_m}} \eta_m(g)\eta_m(h) \le \frac{4}{n_0^{\varepsilon}}.$$

□

We next consider a typical pair $(g,h) \in T_{\eta_m}$. Notice that we can write $\mu * \delta_g(x) = \mu(xg^{-1})$ and $\mu * \delta_h(x) = \mu(xh^{-1})$. Thus, with $k = hg^{-1}$, by definition

$$\sum_{x\in G}(\mu(x) - \mu(xk))^2 = \sum_{x\in G}(\mu(xg^{-1}) - \mu(xh^{-1}))^2 \le \frac{1}{n_0^{1-\varepsilon}} \sum_{x\in G} \mu^2(x). \tag{19}$$

Thus it is natural to introduce the "distance" with respect to μ:

$$d_\mu(g,h) := \sqrt{\frac{\sum_{x\in G}(\mu(xg^{-1}) - \mu(xh^{-1}))^2}{\sum_{x\in G}\mu^2(x)}}.$$

Thus (g,h) is typical iff

$$d_\mu(g,h) \le \frac{1}{n_0^{1/2-\varepsilon/2}}. \tag{20}$$

Using definition, we can show the following elementary properties about d_μ.

Fact 3.2. *For every k we have $d_\mu(k, id_G) = d_\mu(k^{-1}, id_G)$. Furthermore d_μ is right-invariant, symmetric, and satisfies the triangle inequality.*

For the remaining part of this section we will continue to understand further properties of d_μ given the structure of $\mathrm{supp}(\mu)$ obtained from Theorem 2.7. As μ is fixed, allow us to drop the subscript μ in $d_\mu(.)$ for convenience. We first show that the set of k of small distance to id_G can be covered efficiently.

Claim 3.3. *For δ sufficiently small depending on c (from Lemma 2.3), there exists a collection of $O(c^{-O(1)})$-left translations of HP^2 which contains all k with $d(k, id_G) \le \delta$.*

Proof. (of Claim 3.3) Let x_0HP be the coset nilprogression obtained from Theorem 2.7. By assumption $d(k^{-1}, id_G) = d(k, id_G) \le \delta$,

$$\sum_{x\in x_0HP}(\mu(x)-\mu(xk^{-1}))^2\leq\sum_{x\in G}(\mu(x)-\mu(xk^{-1}))^2\leq\delta^2\|\mu\|_2^2. \tag{21}$$

As $(\mu(x)-\mu(xk^{-1}))^2\geq\frac{1}{2}\mu^2(x)-\mu^2(xk^{-1})$, it follows that

$$\sum_{x\in x_0HP}\mu^2(xk^{-1})\geq\frac{1}{2}\sum_{x\in x_0HP}\mu^2(x)-\delta^2\|\mu\|_2^2.$$

Thus if we choose $\delta\leq\delta_0$ with sufficiently small δ_0 depending on c,

$$\sum_{x\in x_0HP}\mu^2(xk^{-1})\geq K^{-O(1)}\|\mu\|_2^2. \tag{22}$$

We now consider a maximal collection of *disjoint* left translations of the form

$$\Big\{k_iHP, 0\leq i\leq N\Big\},\ \text{where } k_0=id_G \text{ and } d(k_i,id_G)\leq\delta, i\geq 1. \tag{23}$$

By disjointness (and as $P=P^{-1}$ and $HP=PH$),

$$x_0HPk_i^{-1}\cap x_0HPk_j^{-1}=\emptyset.$$

By (22) we must have

$$N=K^{O(1)}.$$

By the maximality assumption, for any k with $d(k,id_G)\leq\delta$ there exists k_i such that $kHP\cap k_iHP\neq\emptyset$, thus

$$k\in k_iHP(HP)^{-1}\subset k_iHP^2.$$

□

Let $C_0=N+1=O(K^{O(1)})=O(c^{-O(1)})$ be the constant obtained from the proof of Claim 3.3. We can always assume $C_0\geq 2$. As we can always extend a maximal collection of disjoint translations of form (23) with respect to δ_1 (which plays the role of δ in Claim 3.3) to a maximal one with respect to $\delta_2\geq\delta_1$, and because we have seen from the proof of Claim 3.3 that each such maximal collection has at most C_0 members as long as $\delta\leq\delta_0$, there exists an integer $l=O_K(1)$ such that

$$\mathcal{C}_{\delta_0/C_0^{l+1}}=\mathcal{C}_{\delta_0/C_0^{l-1}}, \tag{24}$$

where $\mathcal{C}_{\delta_0/C_0^{l+1}}$ and $\mathcal{C}_{\delta_0/C_0^{l-1}}$ are such two maximal collections with respect to $\delta_1 = \delta_0/C_0^{l+1}$ and $\delta_2 = \delta_0/C_0^{l-1}$. Let $\{k_{l-1,1}, \dots, k_{l-1,r}\} = \{k_{l+1,1}, \dots, k_{l+1,r}\}$ be the representatives with respect to $\mathcal{C}_{\delta_0/C_0^{l+1}}$ (and equivalently, with respect to $\mathcal{C}_{\delta_0/C_0^{l-1}}$, and hence also with respect to $\mathcal{C}_{\delta_0/C_0^{l}}$) where $r \le C_0$.

In the next step we define T to be the collection of the left cosets $k_{l,i}\langle HP\rangle$. Because of our definition (23) that every maximal collection contains HP, T contains the coset $t_{id_G} = \langle HP\rangle$.

One can put a "distance" d_T on the coset elements of T as

$$d_T(x\langle HP\rangle, y\langle HP\rangle) := \inf_{g\in G}\Big\{d(g, id_G) : gx\langle HP\rangle = y\langle HP\rangle\Big\}.$$

We remark that if $d_T(.)$ is well defined on the coset elements of T then it does not depend on the representatives and it is symmetric. To show that it is well defined, for any vertex pair $(t,t') = (k_{l,i}\langle HP\rangle, k_{l,j}\langle HP\rangle)$ in T, because $d(k_{l,i}, id_G)$ and $d(k_{l,j}, id_G)$ are both finite, $d_T(k_{l,i}\langle HP\rangle, id_G\langle HP\rangle)$ and $d_T(id_G\langle HP\rangle, k_{l,j}\langle HP\rangle)$ are finite, and so is $d_T(k_{l,i}\langle HP\rangle, k_{l,j}\langle HP\rangle)$ by the triangle inequality with respect to $d(.)$. More precisely,

$$\begin{aligned} d_T(t,t') = d_T(k_{l,i}\langle HP\rangle, k_{l,j}\langle HP\rangle) &\le d_T(k_{l,i}\langle HP\rangle, id_G\langle HP\rangle) + d_T(id_G\langle HP\rangle, k_{l,j}\langle HP\rangle) \\ &\le d(k_{l,i}, id_G) + d(k_{l,j}, id_G) \le 2\delta/C_0^{l+1} \le \delta/C_0^l. \end{aligned} \tag{25}$$

Next we consider the weighted complete graph G on T with weights $w(f) = d_T(t,t')$ on any edge $f = (t,t') \in \binom{T}{2}$.

Claim 3.4. *There exists a spanning tree F of G with the following properties*

(1) for each pair $(t,t') \in \binom{T}{2}$, each weight of the edges on the tree path connecting t to t' is at most $d_T(t,t')$;

(2) one can also choose corresponding coset representatives x_t for each $t \in T$ such that as long as (t,t') is an edge of F

$$d_T(t,t') = d(x_t, x_{t'});$$

(3) furthermore, for any $(t,t') \in \binom{T}{2}$

$$d_T(t,t') \asymp_K d(x_t, x_{t'}). \tag{26}$$

The proof of this claim follows [17, Lemma 3.2], we present it here for the reader's convenience.

Proof. (of Claim 3.4) We construct the tree and the coset representatives by a simple greedy algorithm starting from step 0 with $F_0 = \{id_G\}$. Assume that at step i we already obtain a

subtree F_i with the coset representative x_t for each $t \in F_i$, we then find an edge connecting F_i to $T\backslash V(F_i)$ of least weight, say $e = (t, t')$. It is clear that for any $t'' \in F_i$

$$d_T(t, t') \leq d_T(t'', t'). \tag{27}$$

Let g be an element from G such that $d(g, id_G) = d_T(t, t')$ by the definition of $d_T(.)$. In the next step set $x_{t'} := gx_t$ and $F_{i+1} := F_i \cup \{e\}$, we continue the process until the last vertex.

The first claim then follows from (27) and the way F was constructed. The second claim also follows because x_t do not change along the construction process. For the third claim, first recall that $|V(T)| = O_K(1)$. Assume that $t_0 = t, t_1, \dots, t_{j-1}, t_j = t'$ is the F-path connecting t to t'. By the triangle inequality

$$\begin{aligned} d_T(t, t') &\leq \sum_{t_0=t,t_1,\dots,t_{j-1},t_j=t',(t_i,t_{i+1})\in F} d_T(t_i, t_{i+1}) \\ &= \sum_{t_0=t,t_1,\dots,t_{j-1},t_j=t',(t_i,t_{i+1})\in F} d(x_{t_i}, x_{t_{i+1}}) \leq |V(T)| d(x_t, x_{t'}), \end{aligned}$$

where in the last estimate we used the first claim (1). For the other direction, again by (1) and by the triangle inequality

$$\begin{aligned} d_T(t, t') &\geq \frac{1}{j} \sum_{t_0=t,t_1,\dots,t_{j-1},t_j=t',(t_i,t_{i+1})\in F} d_T(t_i, t_{i+1}) \\ &= \frac{1}{j} \sum_{t_0=t,t_1,\dots,t_{j-1},t_j=t',(t_i,t_{i+1})\in F} d(x_{t_i}, x_{t_{i+1}}) \geq \frac{1}{|V(T)|} d(x_t, x_{t'}). \end{aligned}$$

□

Set

$$X := \{x_t : t \in T\}.$$

Then $|X| \leq C_0 = O(K^{O(1)})$ and the cosets $x\langle HP\rangle, x \in X$ are disjoint. Furthermore, assume that x comes from the coset $t = k_{l,i}\langle HP\rangle$, then

$$d(x, id_G) \leq |V(T)| d_T(t, t_{id_G}) \leq |V(T)| d(k_{l,i}, id_G) \leq |V(T)| \delta_0 / C_0^{l+1} \leq \delta_0 / C_0^l, \tag{28}$$

where we recall that $k_{l,i}$ is one of the representatives of $\mathcal{C}_{\delta_0/C_0^{l+1}}$.

By Claim 3.3 and by (24) we have

$$x \in k_{l,i} HP^2.$$

In other words,

$$k_{l,i} \in xHP^2.$$

We also notice that this holds for any representative $k_{l,i}$ where $x \in k_{l,i}\langle HP\rangle$. Thus, again by Claim 3.3 and (24)

$$\left\{g : d(g, id_G) \le \delta/C_0^{l-1}\right\} \subset \cup_{i=1}^{r} k_{l,i} HP^2 \subset \cup_{x\in X}^{*} xHP^4, \tag{29}$$

where the disjointness comes from the mentioned fact that $x\langle HP\rangle, x \in X$, are disjoint.

We now establish the connection between $\|.\|_{HP,X}$ and $d(.)$.

Lemma 3.5. *As long as $d(g, id_G) \le \delta_0/C_0^l$, we have*

$$\|g\|_{HP,X} \ll d^{1-O(\varepsilon)}(g, id_G).$$

Proof. (of Lemma 3.5) Consider any g with $d(g, id_G) \le \delta_0/C_0^l$. Then for any $x \in X$,

$$d(gx, id_G) \le d(gx, x) + d(x, id_G) = d(g, id_G) + d(x, id_G) \le \delta_0/C_0^l + \delta_0/C_0^l \le \delta_0/C_0^{l-1},$$

where we used (28).

Thus, by (29), $gx \in x'HP^4$ for some $x' \in X$. Write

$$gx = x'h, \text{ for some } h \in HP^4.$$

Note that by the definition of x, x'

$$d(x, x') \le |V(T)| d(x\langle HP\rangle, x'\langle HP\rangle) \le |V(T)| d(g, id_G).$$

Thus

$$d(x', x'h) = d(x', gx) \le d(x', x) + d(x, gx) \le (|V(T)| + 1) d(g, id_G).$$

Again by right invariance and by the triangle inequality

$$d(id_G, h^q) = d(x', x'h^q) \le q(|V(T)| + 1) d(g, id_G).$$

Let q be the largest power of 2 that is smaller than $\delta_0/C_0^l d(g, id_G)$

$$\frac{\delta_0}{2C_0^l d(g, id_G)} \leq q = 2^k \leq \frac{\delta_0}{C_0^l d(g, id_G)}. \tag{30}$$

Thus $d(id_G, h^q) < \delta_0/C_0^{l-1}$, and so by (29) and by the fact that $h \in HP^4$

$$h^q \in (\cup^*_{x\in X} xHP^4) \cap \langle HP\rangle.$$

Because $id_G \in X$ and the cosets $x\langle HP\rangle, x \in X$, are all disjoint, we obtain

$$h^q \in HP^4.$$

By the properness of HP, after iterating the third conclusion of Claim 2.8 k times, we obtain that $h \in HP_{(1+1/D)^k/2^k} \subset HP_{1/q^{1-O(\varepsilon)}}$ as D was chosen to be larger than $1/\varepsilon$. Hence,

$$\|h\|_{HP} = O(\frac{1}{q^{1-O(\varepsilon)}}) \ll d^{1-O(\varepsilon)}(g, id_G). \tag{31}$$

Thus we have

$$\|g\|_{HP,X} \ll d^{1-O(\varepsilon)}(g, id_G).$$

To complete the proof, we note that the map $x \to x'$ above depends on g and it is one-to-one because the representatives x come from different cosets of $\langle HP\rangle$.

□

4. Proof of Theorem 1.14: third step

We show the following form of Theorem 1.14.

Theorem 4.1 (Structures for a_m's)**.** *There exists a coset nilprogression HP in $O(1)$-normal form of small rank and step with $|HP| = O(\rho^{-1})$, and a finite set X of cardinality $O(1)$ such that or all $1 \leq m \leq n^{1-\varepsilon}$,*

$$\|a_m {a'_m}^{-1}\|^2_{HP,X} \leq \frac{1}{n_0^{1-O(\varepsilon)}}, \text{ for all } a_m, a'_m \in A_{j_0-m}.$$

Proof. (of Theorem 4.1) First observe that

$$\eta_m = \mu_{j_0} * \cdots * \mu_{j_0-m} := \eta_{m-1} * \mu_{j_0-m}.$$

Claim 4.2. *For any $a, a' \in A_{j_0-m}$ there is a typical pair with respect to η_m of the form (ga, ha'), where (g, h) is also a typical pair with respect to η_{m-1}.*

Proof. (of Claim 4.2) First of all, if we look at the typical pairs of η_{m-1}, then by Claim 3.1

$$\sum_{(g,h)\notin T_{\eta_{m-1}}} \eta_{m-1}(g)\eta_{m-1}(h) \leq \frac{4}{n_0^{\varepsilon}}.$$

Let $T_{a,a'}$ be the collection of pairs of words (g', h') in $\mathrm{supp}(\eta_m) \times \mathrm{supp}(\eta_m)$ of the form (ga, ha') where (g, h) forms a typical pair with respect to η_{m-1}. Then by (1)

$$\sum_{(g',h')\in T_{a,a'}} \eta_m(g')\eta_m(h') \geq p_0^2 \sum_{(g,h)\in T_{\eta_{m-1}}} \eta_{m-1}(g)\eta_{m-1}(h) > p_0^2/2.$$

On the other hand, by Claim 3.1 applied to η_m

$$\sum_{(g',h')\notin T_{\eta_m}} \eta_m(g')\eta_m(h') \leq \frac{4}{n_0^{\varepsilon}}.$$

But as $p_0 \geq 1/n^{\varepsilon^3} \geq (8/n_0^{\varepsilon})^{1/2}$, we have

$$T_{a,a'} \cap T_\nu \neq \emptyset.$$

So there is a typical pair (g, h) with respect to η_m satisfying the conclusion. □

Let HP be the coset nilprogression obtained from Theorem 2.7, for which by (7)

$$|HP| = O(\rho^{-1}).$$

For any $1 \leq m \leq n^{1-\varepsilon}$, and for any $a, a' \in A_{j_0-m}$ consider a ν_{m-1}-typical pair (g, h) so that (ga, ha') is also a ν_m-typical pair. By right invariance,

$$d(gaa'^{-1}, h) = d(ga, ha') \ll n_0^{-1/2+\varepsilon/2}.$$

Thus

$$d(a'a^{-1}g^{-1}h, id_G) \ll n_0^{-1/2+\varepsilon/2}.$$

Furthermore, as (g,h) is ν_{m-1}-typical

$$d(g^{-1}h, id_G) \ll n_0^{-1/2+\varepsilon/2}.$$

By the triangle inequality,

$$d(a'a^{-1}, id_G) = d(a'a^{-1}g^{-1}h, g^{-1}h) \leq d(a'a^{-1}g^{-1}h, id_G) + d(id_G, g^{-1}h) \ll n_0^{-1/2+\varepsilon/2}. \tag{32}$$

The proof of Theorem 4.1 is then complete by Lemma 3.5. □

We remark that the use of triangle inequality to obtain (32) as above is rather wasteful. We suspect the following.

Conjecture 4.3. *Assume that μ_i are as in Theorem 1.14 with $id_G \in A_i$ such that $\rho(\mu_1, \dots, \mu_n) \geq n^{-O(1)}$. Then there exist consecutive indices $i_0, \dots, i_0 + n^{1-\varepsilon}$ such that*

$$\sum_{i_0 \leq i \leq i_0 + n^{1-\varepsilon}} \sum_{a_i \in A_i} \|a_i\|_{HP,X}^2 \ll 1.$$

This bound, if true, would be a non-abelian analog of [18, Equation 7.9] and it would directly yield the second conclusion of Theorem 1.5.

5. The Erdős-Littlewood-Offord bound in non-abelian groups

To prove Theorem 1.5 we will follow the proof of Theorem 1.14. Assume for contradiction that for some sufficiently large constant

$$\|\mu_n * \cdots * \mu_1\|_\infty \geq C_0 \max\{\frac{1}{s}, \frac{1}{n^{1/2-\delta}}\}.$$

Without loss of generality (by passing to $n/2$ consecutive μ_i, see also (7)) we can assume $\|\mu_n * \cdots * \mu_1\|_2^2 \geq C_0 \max\{s^{-1}, n^{-1/2+\delta}\}$. We will choose C_0 to be larger than any other implied constants in the sequel.

Argue as in Section 2, by (9), there exists $0 \leq k \leq n^{1-\varepsilon}$ such that

$$\frac{\|\mu_{[i_0+l_0-(j+1)n^{1-\varepsilon}+k, i_0+2l_0]}\|_2}{\|\mu_{[i_0+l_0-(j+1)n^{1-\varepsilon}+k+1, i_0+2l_0]}\|_2} \geq 1 - \frac{1}{n^{1-\varepsilon}}.$$

Set $j_0 = i_0 + l_0 - (j+1)n^{1-\varepsilon} + k + 1$ and $l_0^* = i_0 + 2l_0 - j_0$. We obtain the following analog of Lemma 2.3.

Lemma 5.1. *The exist j_0, l_0^* with $l_0^* \geq n^{1-\varepsilon}$ such that*

$$\|\mu_{[j_0-l_0^*,j_0+l_0^*]}\|_2 \geq c \max\Big\{\|\mu_{[j_0-l_0^*,j_0-1]}\|_2, \|\mu_{[j_0,j_0+l_0^*]}\|_2\Big\} \tag{33}$$

and

$$|\mu_{[j_0-1,j_0+l_0^*]}\|_2 \geq (1-n^{1-\varepsilon})\|\mu_{[j_0,j_0+l_0^*]}\|_2. \tag{34}$$

Set $\mu := \mu_{[j_0,j_0+l_0^*]}, \nu := \mu_{[j_0-l_0^*,j_0-1]}$, we follow Section 2 to obtain Theorem 2.7 for μ, ν.

In the next step, let

$$n_0 = n^{1-\varepsilon} \text{ and } \eta = \mu_{j_0-1}.$$

Note that in contrast to Section 3 and Section 4, our n_0 here is large and we will only focus on one special η (instead of many η_m). By (34) we have

$$\|\mu * \eta\|_2 \geq (1-1/n_0)\|\mu\|_2.$$

By the argument of Section 3, especially by combining Claim 3.1 (for $\eta_m = \mu_{j_0-1}$), equation (20) and Lemma 3.5, we obtain the following analog of Theorem 4.1.

Theorem 5.2. *There exists a coset nilprogression HP in $O(1)$-normal form of small rank and step with $|HP| = O(\frac{1}{C_0}\min\{s, n^{1/2-\delta}\})$, and a finite set X of cardinality $O(1)$ and a distribution μ_{j_0-1} whose support contains a pair $\{a, a'\}$ such that aa'^{-1} has order at least s and*

$$\|aa'^{-1}\|_{HP,X}^2 \leq n_0^{-1+O(\varepsilon)} < n^{-1+O(\varepsilon)}.$$

Now consider the bound $\|aa'^{-1}\|_{HP,X} \leq n^{-1/2+C\varepsilon}$ for some absolute constant C. If we choose ε so that $\delta > C\varepsilon$, then

$$|HP| = O(\min\{s, n^{-1/2+\delta}\}) = O(n^{1/2-\delta}) < n^{1/2-C\varepsilon}.$$

Thus, the bound $\|aa'^{-1}\|_{HP,X} \leq n^{-1/2+C\varepsilon}$ forces p to be in H for any representation of the form $xp\sigma(x)^{-1}$ of aa'^{-1} with $p \in HP$. In other words, for all $x \in X$

$$aa'^{-1} \in xH\sigma(x)^{-1}.$$

Replace $x = \sigma(x)$ and iterate the relation d times where d is the order of σ in $Sym(X)$. After multiplying the obtained identities, we have

$$(aa'^{-1})^d \in xHx^{-1}.$$

However, this would imply that the order k of aa'^{-1} is at most

$$k \leq d|H| = O(|HP|) = O(\frac{1}{C_0}s) < s,$$

where C_0 was chosen sufficiently large. This contradicts with our assumption that aa'^{-1} must have order at least s.

6. The Sárközy-Szemerédi's bound in non-abelian groups

We prove Theorem 1.6. Assume otherwise, then again by passing to $n/2$ consecutive μ_i we can assume $\|\mu_n * \cdots * \mu_1\|_2^2 \gg n^{-1+\delta}$. By Theorem 4.1, with $\varepsilon = \delta/2$, there exists a coset nilprogression HP with the following properties

(1) P has rank and step $r, s = O(1)$ and $|HP| = O(n^{1-\delta})$;

(2) There is a finite set X of cardinality $|X| = O(1)$, and consecutive indices $i_0, \ldots, i_0 + n'$ with $n' = n^{1-\varepsilon}$ such that

$$\sup_{i_0 \leq i \leq i_0+n'} \|a_i a_i'^{-1}\|_{HP,X} < 1.$$

More specifically, each element $a_i a_i'^{-1}, i_0 \leq i \leq i_0 + n^{1-\varepsilon}$, can be written as $xh(x')^{-1}$ for some $x, x' \in X$ and $h \in HP$. However, this is impossible when $\varepsilon = \delta/2$ because the $a_i a_i'^{-1}$ are distinct and there are only $|X|^2|HP| = O(n^{1-\delta})$ ways to choose for the values of $a_i a_i'^{-1}$ from the set $XHPX^{-1}$.

7. Proof of Theorem 1.15

Assume otherwise that for some positive constant A

$$\rho = \sup_{g \in \mathbf{Sl}_2(\mathbf{R})} \mathbf{P}(g_n \ldots g_1 = g) \geq n^{-A}.$$

By Theorem 1.14, there exists a nilprogression HP with size $|HP| = O(n^A)$ and there exist a finite set X of cardinality $|X| = O(1)$ and indices $i_0, \ldots, i_0 + n'$ with $n' = n^{1-O(\varepsilon)}$ such that the following holds: for each $a \in \mathrm{supp}(\mu_i), i_0 \leq i \leq i_0 + n'$ there exists a permutation $\sigma_a \in Sym(X)$ such that for all $x \in X$,

$$\begin{pmatrix} E + \lambda a & -1 \\ 1 & 0 \end{pmatrix} \in xHP(\sigma_a(x))^{-1}.$$

By our assumption, among these n' consecutive μ_i, there exists one whose support contains $a, -a$ with $a > \gamma$. We will be focusing on these two elements. For short, write

$$g_1 := \begin{pmatrix} E+\lambda a & -1 \\ 1 & 0 \end{pmatrix} \text{ and } g_2 := \begin{pmatrix} E-\lambda a & -1 \\ 1 & 0 \end{pmatrix}.$$

By definition, for any integer k the ball $B_k(g_1, g_2)$ which consists of words of length at most k in $g_1^{\pm 1}, g_2^{\pm 1}$ has size

$$|B_k(g_1, g_2)| \leq |XHP^kX| = O(k^{O(1)}|HP|) = O(k^{O(1)}n^C), \tag{35}$$

where we used (6) in the estimate of HP^k.

On the other hand, $(g_1)^{-1} = \begin{pmatrix} 0 & 1 \\ -1 & E+\lambda a \end{pmatrix}$ and $(g_2)^{-1} = \begin{pmatrix} 0 & 1 \\ -1 & E-\lambda a \end{pmatrix}$. So

$$h_1 = g_1(g_2)^{-1} = \begin{pmatrix} 1 & 2\lambda a \\ 0 & 1 \end{pmatrix} \text{ and } h_2 = (g_1)^{-1}g_2 = \begin{pmatrix} 1 & 0 \\ 2\lambda a & 1 \end{pmatrix}.$$

Choose $k_0 = \lceil 1/2\lambda \rceil$ so that $2k_0\lambda \geq 1$, and consider

$$h_1' := h_1^{k_0} = \begin{pmatrix} 1 & 2k_0\lambda a \\ 0 & 1 \end{pmatrix} \text{ and } h_2' := h_2^{k_0} = \begin{pmatrix} 1 & 0 \\ 2k_0\lambda a & 1 \end{pmatrix}.$$

We next use the following lemma.

Lemma 7.1. [1] *If $\mu \in \mathbf{R}$ with $|\mu| \geq 2$ then the group generated by the matrices $\begin{pmatrix} 1 & \mu \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 1 & 0 \\ \mu & 1 \end{pmatrix}$ is free.*

Thus by Lemma 7.1, for any k

$$|B_k(g_1, g_2)| \geq |B_{k/2k_0}(h_1', h_2')| \geq 2^{k/2k_0}.$$

However this would contradict with the polynomial bound (35).

References

[1] J. Brenner, *Quelques groupes tibris de matrices*, CR Acad. Sc, Paris 241 (1955), 1689-1691.

[2] E. Breuillard and M. Tointon, *Nilprogressions and groups with moderate growth*, Adv. Math. 289 (2016), 1008-1055.

[3] E. Breuillard, B. Green, R. Guralnick, and T. Tao, *Expansion in finite simple groups of Lie type*, Journal of the European Mathematical Society, Vol. 17 (2015), Issue 6, 1367-1434.

[4] E. Breuillard, B. Green, and T. Tao, *The structure of approximate groups*, Publications Mathematiques Institut de Hautes Etudes Scientifiques, 116 (2012), 115-221.

[5] N. Coulhon, L. Saloff-Coste, and N. Varopoulos, *Analysis and Geometry on groups*, Cambridge University Press, 1992.

[6] P. Erdős, *On a lemma of Littlewood and Offord*, Bull. Amer. Math. Soc. 51 (1945), 898-902.

[7] P. Erdős and L. Moser, *Elementary Problems and Solutions: Solutions: E736*. American Mathematical Monthly, 54 (1947), no. 4, 229-230.
[8] G. Halász, Estimates for the concentration function of combinatorial number theory and probability, Period. Math. Hungar. 8 (1977), no. 3-4, 197-211.
[9] J. E. Littlewood and A. C. Offord, *On the number of real roots of a random algebraic equation*. III. Rec. Math. Mat. Sbornik N.S. 12 , (1943). 277–286.
[10] H. Nguyen, A new approach to an old problem of Erdős and Moser, Journal of Combinatorial Theory, Series A 119 (2012) 977-993
[11] H. Nguyen and V. Vu, *Optimal Littlewood-Offord theorems*, Advances in Mathematics, Vol. 226, 6 (2011), 5298-5319.
[12] H. Nguyen and V. Vu, *Small probability, inverse theorems, and applications*, Paul Erdos' 100th anniversary, Bolyai Society Mathematical Studies, Vol. 25 (2013)).
[13] A. Sárközy and E. Szemerédi, *Uber ein Problem von Erdős und Moser*, Acta Arithmetica, 11 (1965) 205-208.
[14] R. Stanley, Weyl groups, the hard Lefschetz theorem, and the Sperner property, SIAM J. Algebraic Discrete Methods 1 (1980), no. 2, 168-184.
[15] T. Pham and V. Vu, *Non-abelian Littlewood-Offord inequalities*, Adv. Math. 302 (2016), 1233-1250.
[16] T. Tao, *Product set estimates for non-commutative groups*, Combinatorica, 28 (2008), 547-594.
[17] T. Tao, *Inverse theorems for sets and measures of polynomial growth*, to appear, The Quarterly Journal of Mathematics, `arxiv.org/abs/1507.01276`.
[18] T. Tao and V. Vu, *On the singularity probability of random Bernoulli matrices*, Journal of the A. M. S 20 (2007), 603-673.
[19] T. Tao and V. Vu, *From the Littlewood-Offord problem to the circular law: universality of the spectral distribution of random matrices*, Bulletin of the American Mathematical Society, **46** (2009), 377-396.
[20] T. Tao and V. Vu, *Inverse Littlewood-Offord theorems and the condition number of random matrices*, Annals of Mathematics (2), 169 (2009), no 2, 595-632.
[21] T. Tao and V. Vu, *A sharp inverse Littlewood-Offord theorem*, Random Structures Algorithms 37 (2010), no. 4, 525-539.
[22] M. Tointon, *Freiman's theorem in an arbitrary nilpotent group*, Proceedings London Mathematical Society (3) 109 (2014) 318-352.

Department of Mathematics, The Ohio State University, Columbus, OH 43210, USA

E-mail address: `nguyen.1261@math.osu.edu`